

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-154164
(43)Date of publication of application : 08.06.1999

(51)Int.Cl.

G06F 17/30

(21)Application number : 09-337713

(71)Applicant : HITACHI LTD

(22)Date of filing : 21.11.1997

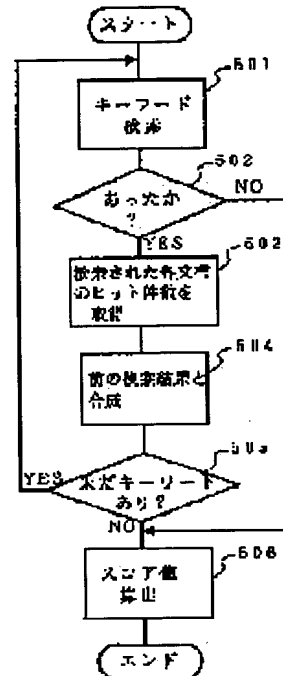
(72)Inventor : HOSHI YUKIO

(54) ADAPTABILITY CALCULATING METHOD IN WHOLE SENTENCE SEARCH PROCESSING AND STORAGE MEDIUM STORING PROGRAM RELATED TO THE SAME

(57)Abstract:

PROBLEM TO BE SOLVED: To calculate adaptability adjusted to the sense of a user by using a value obtained by adding the number of hits for respective search terms when an AND condition is designated a retrieval condition.

SOLUTION: One search term is taken out and an index is searched with the term as a keyword (501). The number of hits of respective documents obtained from the result (503). The result of the search term till previous time and the result of the search term of this time are put together (504). When the term to be search still remains (505), search is repeated and the new result is put together (504). When the AND condition is instructed as the search condition, the result till previous time and the result of this time are AND-operated and the new result is generated/held. The number of hits till previous time and the number of hits of this time are added for the respective documents and the number of hits is obtained. When all the search terms are terminated, a score value is calculated based on the number of hits (506).



LEGAL STATUS

[Date of request for examination] 21.08.2000
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-154164

(43) 公開日 平成11年(1999) 6月8日

(51) Int.Cl.⁶

G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/403

15/40

15/403

3 4 0 C

3 7 0 A

3 7 0 Z

審査請求 未請求 請求項の数 2 F D (全 7 頁)

(21) 出願番号

特願平9-337713

(22) 出願日

平成9年(1997)11月21日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 星 幸雄

神奈川県横浜市戸塚区戸塚町5030番地 株

式会社日立製作所ソフトウェア開発本部内

(74) 代理人 弁理士 矢島 保夫

(54) 【発明の名称】 全文検索処理における適合度算出方法および該方法に係るプログラムを格納した記憶媒体

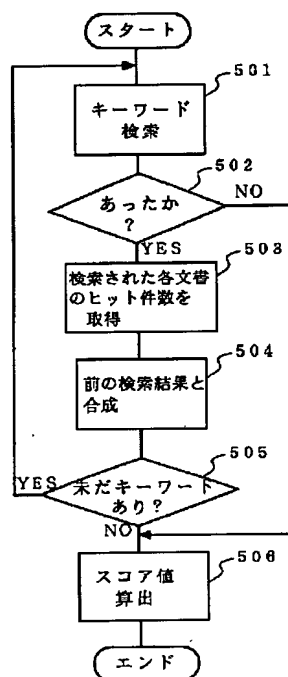
(57) 【要約】

(修正有)

【課題】 Web ページの全テキストから所望の文字列を検索したり、複数の文書の全テキストから所望の文字列を検索する全文検索システムにおいて、AND 検索でも OR 検索でもユーザの感覚に合った使い勝手のよいテキスト検索方法および該方法に係るプログラムを格納した記憶媒体を提供する。

【解決手段】 各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得し、検索条件として AND 条件が指定されていた場合は、取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とし、求めた各文書のヒット件数から、各文書のスコア値を算出する。

検索とスコア値算出処理のフローチャート



【特許請求の範囲】

【請求項1】 指定された複数の検索タームおよび検索条件でテキスト文書を検索し、検索結果を返す全文検索処理における適合度算出方法であって、

各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得する検索ステップと、

検索条件としてAND条件が指定されていた場合、前記検索ステップにより取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とする検索結果合成ステップと、

前記検索結果合成ステップにより得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出ステップとを備えたことを特徴とする全文検索処理における適合度算出方法。

【請求項2】 指定された複数の検索タームおよび検索条件でテキスト文書を検索し、検索結果を返す全文検索処理における適合度算出方法に係るプログラムを格納した記憶媒体であって、

該プログラムは、

各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得する検索ステップと、

検索条件としてAND条件が指定されていた場合、前記検索ステップにより取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とする検索結果合成ステップと、

前記検索結果合成ステップにより得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出ステップとを備えたことを特徴とする全文検索処理における適合度算出方法に係るプログラムを格納した記憶媒体。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】 本発明は、全文検索処理における適合度算出方法および該方法に係るプログラムを格納した記憶媒体に関する。

【0002】

【従来の技術】 近年、インターネットのWWW (World Wide Web) の検索サービス (サーチ・エンジン) では、登録したWebページの全テキストから所望の文字列を検索できる検索システムが用いられているものがある。また、各種の文書検索システムでも、各文書の全テキストから所望の文字列を検索するものが知られている。そのような全文検索システムでは、いわゆるランキング検

索を行なうことができるものがある。

【0003】 ランキング検索とは、文書検索時、複数の文書がヒットした場合、それぞれの文書に対し、得点を付与しその得点により文書のソートを行なう機能である。例えば、文字列Xを含む文書を検索して、文書Aではヒット件数が5件 (すなわち、文字列Xが文書Aに5ヶ所出現) で、文書Bではヒット件数が8件 (すなわち、文字列Xが文書Bに8ヶ所出現) であったとする。この場合、ランキング検索機能により、これらのヒット件数のうちの最大値を100とした各文書のスコア値 (適合度)、すなわち文書Bのスコア値が100、文書Aのスコア値が $5 \div 8 \times 100 \approx 63$ を、それぞれ表示することができる。また、検索結果を表示するときに、スコア値の大きい順にソートし、文書Bを先に、文書Aを後に、表示することができる。スコア値を表示することにより、ユーザは、検索された文書Aと文書Bとでは、文書Bの方が検索文字列Xを多く含んでおり、文書Aは文書Bの63%程度の出現率で検索文字列Xを含んでいることが分かる。また、スコア値の大きい順に表示されるので、より多くの検索文字列を含む文書が先に表示され、ユーザにとっては使い勝手がよくなる。

【0004】 上述したような検索システムでは、通常、検索文字列を複数指定することもできる。複数の検索文字列は、例えば、ANDやORなどの検索条件でつながれる。ANDは複数の検索文字列をすべて含む文書を検索せよという意味であり、ORは複数の検索文字列のうちの少なくとも1つを含む文書を検索せよという意味である。

【0005】

【発明が解決しようとする課題】 ところで、上記ランキング検索機能を使用した検索で、ANDやORなどの検索条件を指定して複数の検索文字列で検索を行なう場合、スコア値の計算は拡張ブーリアン (boolean) に基づいていた。したがって、AND検索ではヒット件数の少ない方からスコア値を計算し、OR検索ではヒット件数の多い方からスコア値を計算していた。例えば、検索文字列XとYとでAND検索を行なう場合、文書Aでは検索文字列Xのヒット件数が10件で検索文字列Yのヒット件数が3件であった場合、拡張ブーリアンによればX AND Yのヒット件数は3件ということになり、このヒット件数3件に基づいて文書Aのスコア値が計算される。しかし、このようなヒット件数の求め方では、一般的には使い勝手がよくない。上記の例では、文字列Xのヒット件数が10件もあるのに、X AND Yのヒット件数は3件とされてしまい、ユーザの感覚に合わないことが多いからである。

【0006】 本発明は、上述の従来形における問題点を鑑み、AND検索でもOR検索でもユーザの感覚に合った使い勝手のよい全文検索処理における適合度算出方法および該方法に係るプログラムを格納した記憶媒体を提

供することを目的とする。

【0007】

【課題を解決するための手段】上記目的を達成するために、請求項1に係る発明は、指定された複数の検索タームおよび検索条件でテキスト文書を検索し、検索結果を返す全文検索処理における適合度算出方法であって、各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得する検索ステップと、検索条件としてAND条件が指定されていた場合、前記検索ステップにより取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とする検索結果合成ステップと、前記検索結果合成ステップにより得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出ステップとを備えたことを特徴とする。

【0008】請求項2に係る発明は、指定された複数の検索タームおよび検索条件でテキスト文書を検索し、検索結果を返す全文検索処理における適合度算出方法に係るプログラムを格納した記憶媒体であって、該プログラムは、各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得する検索ステップと、検索条件としてAND条件が指定されていた場合、前記検索ステップにより取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とする検索結果合成ステップと、前記検索結果合成ステップにより得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出ステップとを備えたことを特徴とする。

【0009】

【発明の実施の形態】以下、図面を使って本発明の実施の形態を詳細に説明する。

【0010】図1は、本発明に係る全文検索処理における適合度算出方法を適用したテキスト検索システムのシステム構成図である。WWWブラウザ101とWWWサーバ103が、イントラネットまたはインターネットなどのネットワーク102に接続されている。WWWサーバ103には、LAN（ローカルエリアネットワーク）104を介して全文検索サーバ105が接続されている。全文検索サーバ105には、テキストデータベース106が接続されている。WWWブラウザ101、WWWサーバ103、および全文検索サーバ105は、具体的には、それぞれの機能を果たす所定のソフトウェアが動作しているパーソナルコンピュータやワークステーションからなる。また、WWWブラウザ101には、ユー

ザが操作するマウスやキーボードなどの入力装置と、各種の情報を表示する表示装置が備えられているものとする。

【0011】WWWブラウザ101は、表示装置に検索画面を表示し、入力装置によりユーザの入力を受け付ける。図2は、検索画面の表示例を示す。ウインドウ200内の表示領域201に、検索分野見出し211、検索ボタン212、検索条件見出し213、ANDボタン214、ORボタン215、および検索ターム入力領域221～223が表示されている。検索分野見出し211には、この検索システムで文書検索する分野が表示されている。この例は新聞記事の文書を検索するシステムであるので、検索分野見出し211には「新聞記事」と表示されている。ユーザは、この検索画面上で、検索ターム入力領域221～223に検索したい文字列を入力し、必要に応じてANDボタン214またはORボタン215をオンしてANDまたはORの検索条件を選択し、検索ボタン212をオン（マウスでクリック）する。これにより、WWWブラウザ101から、ユーザの入力に応じた検索タームおよび検索条件で、検索要求がWWWサーバ103に送信される。

【0012】図2において、検索ターム入力領域は221～223の3つが備えられているが、検索タームを1個だけ入力して検索ボタン212をオンして検索をかけてもよいし、検索タームを2個または3個の複数入力して検索ボタン212をオンして検索をかけてもよい。検索タームが1個のみ入力されているときは、ANDボタン214およびORボタン215によるANDまたはORの検索条件の指定は無視される。検索タームが2個または3個入力されている場合は、ANDボタン214またはORボタン215をオンすることにより、複数の検索タームをAND条件で検索するか、OR条件で検索するかを選択することができる。表示上は、選択されているボタンが黒丸で、選択されていないボタンが白丸で、それぞれ、表示される。図2では、検索タームが2個入力され、ANDボタン214が選択状態で、ORボタン215が非選択状態であるから、この状態で検索ボタン212をオンすると、「マルチメディア」と「インターネット」の両方を含む文書を検索する指示を行なったことになる。

【0013】再び図1に戻って、上記図2の検索画面で入力された検索タームおよび検索条件の検索要求がWWWサーバ103で受信されると、WWWサーバ103は、その要求に応じて全文検索サーバ105に検索を依頼する。全文検索サーバ105に接続されているテキストデータベース106には、検索対象の複数の文書ファイル（この例では新聞記事の文書）、およびそれらの文書ファイルを検索するためのインデックスが格納されている。インデックスの構造および検索方式については図4を参照して後述する。全文検索サーバ105は、WW

Wサーバ103からの依頼に基づいて、テキストデータベース106のインデックスを参照し、指示された検索タームを含む文書を検索し、その結果（文書番号）をWWWサーバ103に返す。また、全文検索サーバ105は、検索結果の文書ごとにスコア値（適合度）を算出し、そのスコア値に基づいて、検索結果である文書番号をソートしてWWWサーバ103に返す。

【0014】WWWサーバ103は、全文検索サーバ105からの検索結果、すなわち検索された文書番号とスコア値（文書番号はスコア値に基づいてソートされている）を受け、検索結果一覧の画面データを作成し、その画面データをWWWブラウザ101に送信する。WWWブラウザ101は、その検索結果一覧の画面データを受信して画面表示し、これによりブラウザ101を操作しているユーザは検索結果一覧を見ることができる。

【0015】図3は、WWWブラウザ101に表示される検索結果一覧の表示画面の例である。ウィンドウ300内の表示領域301に、検索分野見出し311、結果一覧見出し312、検索結果である文書に関する表示321～323、および次結果一覧ボタン313が表示されている。検索分野見出し311には、図2の211と同様に、この検索システムで文書検索する分野が「新聞記事」と表示されている。結果一覧見出し312は、その下側に検索結果の一覧を表示することを示す見出しである。検索結果である文書に関する表示321～323は、文書番号の表示およびスコア値の表示からなり、スコア値が大きい順に並べられて表示される。検索された文書の数が多いときには、次結果一覧ボタン313をオンすることにより、検索結果一覧の次のページが表示される。

【0016】図4は、このテキスト検索システムにおける検索方式であるn-gramインデックス方式の説明図である。この方式は、連続するn文字からなる文字列に対して、それを含む文書の識別子（文書番号）と、その文字列が出現する位置を対応づけるインデックステーブルを生成し、それを基に検索タームを含む文書を検索する全文検索方式である。図4では、1-gramインデックスを使用した例を示す。

【0017】図4において、文書001（文書番号が001の文書）が「…海洋には数万種の微生物が生息するといわれ…」という文章を含んでいるとする。この文章の各文字の出現位置は先頭から何文字目かで表し、例えば、「海」は13文字目、「洋」は14文字目、「に」は15文字目、…である。この文書001を1-gramインデックス401に登録すると、各文字のインデックスに文書番号と出現位置が設定される。例えば、文書001中では「に」が15文字目に出現しているの、「に」のインデックス411には、文書番号「001」と出現位置「15」が設定される。また、文書001中では「微」が21文字目に出現しているの「微」のイ

ンデックス412には、文書番号「001」と出現位置「21」が設定される。以下同様にして、文書001の全文を1-gramインデックス401に登録しておく。また、他の文書についても同様にしてインデックスに登録しておく。例えば、「に」のインデックス411には、文書番号「008」と出現位置「12」および「35」が登録されているので、文書番号008の文書の12文字目および35文字目に「に」が出現することが分かる。

【0018】全文検索サーバ105では、このインデックス401を用いた検索を以下に行なう。まずWWWサーバ103から、検索ターム402として「微生物」が与えられて検索が指示されたとする。この検索ターム402を1文字ずつインデックス401と照合すると、「微」のインデックス412に文書番号「001」の出現位置「21」が登録されており、「生」のインデックス413に文書番号「001」の出現位置「22」が登録されており、「物」のインデックス414に文書番号「001」の出現位置「23」が登録されているので、文書番号001の文書の出現位置21～23の位置に「微生物」という文字列が存在することが分かる。同様に、文書番号056の出現位置5～7の位置に「微生物」という文字列が存在することが分かる。この検索結果は、文書番号と出現位置の羅列403で表される。

【0019】図4に図示した検索結果403は、この検索ターム「微生物」402が、文書番号001の文書の21文字目に出現し、文書番号056の5文字目に出現することを示しているが、1つの文書中に検索タームが複数含まれているときには、インデックスからその出現位置がすべて検索され検索結果403に反映される。例えば、文書番号001の文書の21文字目のほかに、38文字目および45文字目に検索ターム「微生物」402が出現するとすると、検索結果403は、文書番号001のあとに出現位置21、38、45が続くことになる。なお、文書番号とその文書内の出現位置のデータとの組が1つのまとまりと把握され、そのような組データの区切り位置（図4の検索結果403では、出現位置21と文書番号056の境界）は、分離記号などにより明確になっているものとする。

【0020】以上のようにして、検索結果403から、検索タームを含む文書の文書番号と、検索タームがその文書中のどこに出現するかが分かるので、出現位置の数を数えれば、その文書中に検索タームが幾つ含まれるか（すなわち、ヒット件数が）分かる。そこで、全文検索サーバ105では、検索結果から各文書ごとのヒット件数を求め、そのヒット件数に基づいてスコア値を算出し、WWWサーバ103に返す。

【0021】図5は、全文検索サーバ105における検索とスコア値算出処理のフローチャートを示す。ステップ501で、WWWサーバ103からの依頼に含まれる

検索タームを1つ取り出し、その検索タームをキーワードとして、図4で説明したようにインデックスを検索する。これにより、図4の403に示すような検索結果

(その検索タームを含む文書の文書番号と文書中での出現位置との羅列)が得られる。次に、ステップ502で、検索結果があったか否かを判別する。検索結果がなかったときは、ステップ506に分岐する。検索結果があったときは、ステップ503で、その検索結果から各文書のヒット件数を求める。上述したように、ヒット件数は各文書の出現位置の数を数えれば分かる。いま、例えば1個目の検索タームで検索した結果、以下の(検索結果1)のように文書番号とヒット件数が求められたとする。

【0022】(検索結果1)

文書番号002:ヒット件数6

文書番号008:ヒット件数9

文書番号016:ヒット件数14

【0023】次に、ステップ504で、前回までの検索タームの検索結果と今回の検索タームの検索結果とを合成する。始めは1個目の検索タームの検索であり前回までの検索結果はないから、ステップ504では何もせず、上記の検索結果を保持するだけである。次に、ステップ505で、未だ検索すべき検索タームが指定されているか否かを判別する。なければ、1個目の検索タームの検索であったということであるから、ステップ506で保持されている検索結果に基づいて各文書のスコア値を算出し、求めたスコア値に基づいてソートした検索結果と該スコア値とをWWWサーバ103に返す。

【0024】ステップ505で未だ検索すべき検索タームが残っていたら、ステップ501に戻って、残っている検索タームで検索を行なう。この場合、ステップ504では、前回までに行なわれた検索結果が保持されているので、その検索結果に新たな検索結果を合成する。この合成の処理について具体例で詳しく説明する。いま、1個目の検索タームXで検索した結果、上記(検索結果1)が保持されており、次に2個目の検索タームYで検索して下記の(検索結果2)がステップ503で得られたとする。

【0025】(検索結果2)

文書番号005:ヒット件数7

文書番号008:ヒット件数3

文書番号016:ヒット件数21

【0026】検索条件としてAND条件が指示されているとき、ステップ504の合成処理では、前回までの(検索結果1)と今回の(検索結果2)とのANDを取って新たな検索結果を作成保持する。いま説明している検索結果の例では、文書番号008と016が(検索結果1)と(検索結果2)の両方に存在するので、これらが抽出される。また、ヒット件数については、各文書ごとに前回までの(検索結果1)のヒット件数と今回の

(検索結果2)のヒット件数とを加算して新たなヒット件数にする。したがって、新たに生成される検索結果は下記の(検索結果3)になる。

【0027】(検索結果3)

文書番号008:ヒット件数12

文書番号016:ヒット件数35

【0028】文書番号008のヒット件数12は、(検索結果1)の文書番号008のヒット件数9と(検索結果2)の文書番号008のヒット件数3とを加算した値である。また、文書番号016のヒット件数35は、(検索結果1)の文書番号016のヒット件数14と(検索結果2)の文書番号016のヒット件数21とを加算した値である。

【0029】さらに、未だ検索すべき検索タームがある場合は、上記(検索結果3)を前回までの検索結果としてステップ501からの処理を繰り返せばよい。すべての検索タームについて検索が終了したら、ステップ506で、その時点で保持されている検索結果に基づいてスコア値を算出する。スコア値は、従来と同様にヒット件数が最大のものを100として、各文書ごとに求める。

【0030】上記はAND条件が指定された場合について説明したが、OR条件が指定された場合も同様である。AND条件では前回までの検索結果と今回の検索結果との両方に含まれる文書番号を抽出したが、OR条件では前回までの検索結果と今回の検索結果との少なくとも何れか一方に含まれる文書番号を抽出すればよい。また、ヒット件数の求めかたは、AND条件の場合と同様に、各文書のヒット件数を加算して新たなヒット件数とする。

【0031】上記発明の実施の形態によれば、AND条件で複数の検索タームで検索した場合でも、各文書のヒット件数を加算した値に基づいてスコア値を算出しているので、従来の拡張ブーリアンで実施していたヒット件数の少ない方からスコア値を計算する方式に比較して、ユーザにとって、より使い勝手の良い検索システムになる。

【0032】なお、上記実施の形態では、AND条件のときに各文書のヒット件数を加算した値からスコア値を算出しているが、加算する代わりに、各文書のヒット件数のうちの最大値を用いてスコア値を算出してもよい。また、上記実施の形態では1-gramインデックスの例で説明したが、nを2以上としたn-gramインデックスでも同様に適用可能である。

【0033】

【発明の効果】以上説明したように、本発明によれば、AND条件で複数の検索タームを指定して検索する場合に、各文書のヒット件数を各検索タームのヒット件数の和としているので、従来、ある検索タームのヒット件数は多いが別の検索タームのヒット件数が少ないためにスコア値が大きくなり(ヒット件数が少ない方に基づい

てスコア値を算出していたためである)、上位に表示されなくなる文書があるという不都合が解消され、そのような文書でも上位に表示されるようになる。したがって、ユーザにとって使い勝手の良いテキスト検索を行なうことができる。

【図面の簡単な説明】

【図1】 本発明に係るテキスト検索システムのシステム構成図

【図2】 検索画面の表示例を示す図

【図3】 WWWブラウザに表示される検索結果一覧の表

示画面の例を示す図

【図4】 n-gramインデックス方式の説明図

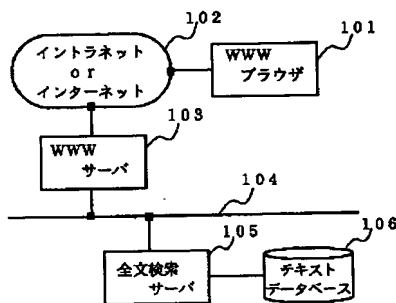
【図5】 全文検索サーバにおける検索とスコア値算出処理のフローチャート図

【符号の説明】

101…WWWブラウザ、102…イントラネットまたはインターネットなどのネットワーク、103…WWWサーバ、104…LAN（ローカルエリアネットワーク）、105…全文検索サーバ、106…テキストデータベース。

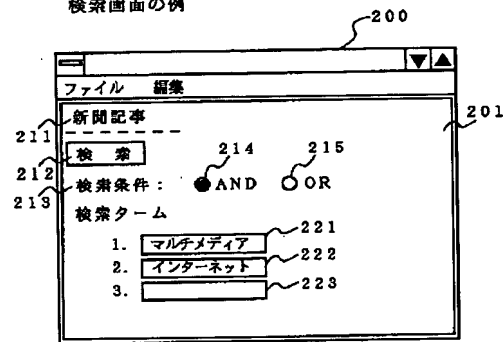
【図1】

本発明に係るテキスト検索システムのシステム構成図



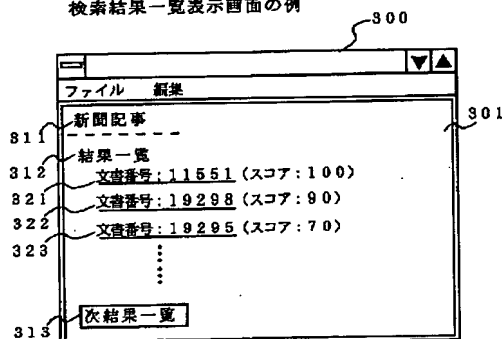
【図2】

検索画面の例



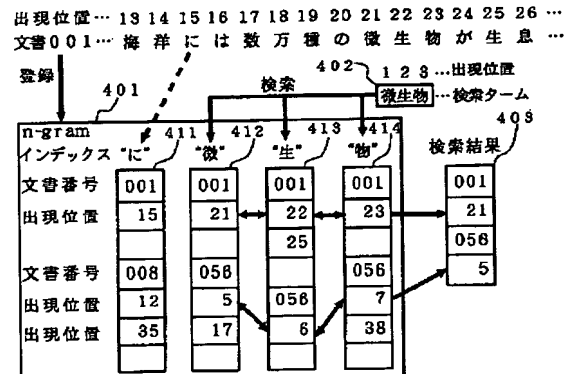
【図3】

検索結果一覧表示画面の例



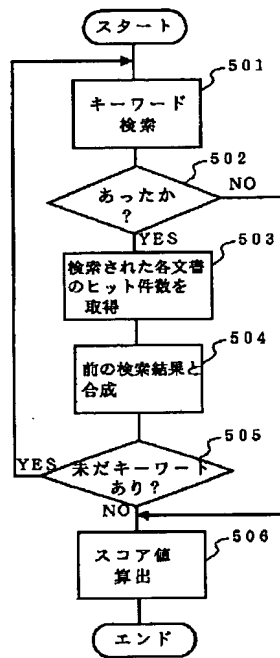
【図4】

n-gram インデックス方式の説明図



【図5】

検索とスコア値算出処理のフローチャート



【公報種別】特許法第 17 条の 2 の規定による補正の掲載
 【部門区分】第 6 部門第 3 区分
 【発行日】平成 13 年 8 月 3 日 (2001. 8. 3)

【公開番号】特開平 11-154164
 【公開日】平成 11 年 6 月 8 日 (1999. 6. 8)
 【年通号数】公開特許公報 11-1542
 【出願番号】特願平 9-337713
 【国際特許分類第 7 版】

G06F 17/30

【F I】

G06F 15/403 340 C
 15/40 370 A
 15/403 370 Z

【手続補正書】

【提出日】平成 12 年 8 月 21 日 (2000. 8. 21)

【手続補正 1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正内容】

【特許請求の範囲】

【請求項 1】指定された複数の検索タームおよび検索条件でテキスト文書を検索し、検索結果を返す全文検索処理における適合度算出方法であって、

各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得する検索ステップと、

検索条件として AND 条件が指定されていた場合、前記検索ステップにより取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とする検索結果合成ステップと、

前記検索結果合成ステップにより得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出ステップとを備えたことを特徴とする全文検索処理における適合度算出方法。

【請求項 2】指定された複数の検索タームおよび検索条件でテキスト文書を検索し、検索結果を返す全文検索処理における適合度算出方法に係るプログラムを格納した記憶媒体であって、

該プログラムは、

各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得する検索ステップと、

検索条件として AND 条件が指定されていた場合、前記検索ステップにより取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とする検索結果合成ステップと、

前記検索結果合成ステップにより得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出ステップとを備えたことを特徴とする全文検索処理における適合度算出方法に係るプログラムを格納した記憶媒体。

【請求項 3】指定された複数の検索タームおよび検索条件でテキスト文書を検索し、検索結果を返す全文検索処理における適合度算出装置であって、

各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得する検索手段と、

検索条件として AND 条件が指定されていた場合、前記検索手段により取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とする検索結果合成手段と、

前記検索結果合成手段により得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出手段とを備えたことを特徴とする全文検索処理における適合度算出装置。

【請求項 4】指定された複数の検索タームでテキスト文書を検索する全文検索処理方法であって、

各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を検索結果として取得する検索ステップと、

前記検索ステップにより取得された各検索タームの検索

結果から、文書の識別子ごとに各検索タームごとのヒット件数を加算して前記識別子に関連付けられた文書のヒット件数を得る検索結果合成ステップと、前記検索結果合成ステップにより得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出ステップとを備えたことを特徴とする全文検索処理方法。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0001

【補正方法】変更

【補正内容】

【0001】

【発明の属する技術分野】本発明は、全文検索処理における適合度算出方法、装置、および該方法に係るプログラムを格納した記憶媒体に関する。

【手続補正 3】

【補正対象書類名】明細書

【補正対象項目名】0006

【補正方法】変更

【補正内容】

【0006】本発明は、上述の従来形における問題点に鑑み、AND検索でもOR検索でもユーザの感覚に合った使い勝手のよい全文検索処理における適合度算出方法、装置、および該方法に係るプログラムを格納した記憶媒体を提供することを目的とする。

【手続補正 4】

【補正対象書類名】明細書

【補正対象項目名】0007

【補正方法】変更

【補正内容】

【0007】

【課題を解決するための手段】上記目的を達成するために、本発明は、指定された複数の検索タームおよび検索条件でテキスト文書を検索し、検索結果を返す全文検索処理における適合度算出方法であって、各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得する検索ステップと、検索条件としてAND条件が指定されていた場合、前記検索ステップにより取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とする検索結果合成ステップと、前記検索結果合成ステップにより得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出ステップとを備えたことを特徴とする。

【手続補正 5】

【補正対象書類名】明細書

【補正対象項目名】0008

【補正方法】変更

【補正内容】

【0008】また本発明は、指定された複数の検索タームおよび検索条件でテキスト文書を検索し、検索結果を返す全文検索処理における適合度算出方法に係るプログラムを格納した記憶媒体であって、該プログラムは、各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得する検索ステップと、検索条件としてAND条件が指定されていた場合、前記検索ステップにより取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とする検索結果合成ステップと、前記検索結果合成ステップにより得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出ステップとを備えたことを特徴とする。また本発明は、指定された複数の検索タームおよび検索条件でテキスト文書を検索し、検索結果を返す全文検索処理における適合度算出装置であって、各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を、検索結果として取得する検索手段と、検索条件としてAND条件が指定されていた場合、前記検索手段により取得された各検索タームの検索結果から、指定されたすべての検索タームを含む文書の識別子を求めるとともに、その文書のヒット件数を、各検索タームごとのヒット件数を加算した値とする検索結果合成手段と、前記検索結果合成手段により得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出手段とを備えたことを特徴とする。さらに本発明は、指定された複数の検索タームでテキスト文書を検索する全文検索処理方法であって、各検索タームごとに検索を実行し、その検索タームを含む文書の識別子、および検索された文書に含まれる検索タームの数を示すヒット件数を検索結果として取得する検索ステップと、前記検索ステップにより取得された各検索タームの検索結果から、文書の識別子ごとに各検索タームごとのヒット件数を加算して前記識別子に関連付けられた文書のヒット件数を得る検索結果合成ステップと、前記検索結果合成ステップにより得られた各文書のヒット件数から、各文書の適合度を算出する適合度算出ステップとを備えたことを特徴とする。